ACCV2002: The 5th Asian Conference on Computer Vision, 23–25 January 2002, Melbourne, Australia.

Estimating the Number of Independent Motions for Multibody Motion Segmentation

Kenichi Kanatani* and Chikara Matsunaga†

* Department of Information Technology, Okayama University, Okayama 700-8530 Japan
†Broadcast Division, FOR-A Co. Ltd., Sakura, Chiba 285-0802 Japan
kanatani@suri.it.okayama-u.ac.jp, matsunaga@for-a.co.jp

Abstract

We study the problem of estimating the number of independent motions for segmentation based on feature tracking. We elucidate the mathematical structure of the problem and present an estimation method using model selection by the geometric AIC, the geometric MDL, and the OIC. We compare their performance using synthetic and real data. We also present techniques for evaluating the reliability of segmentation a posteriori, *using the standard F test and model selection by the geometric AIC and the geometric MDL.*

1. Introduction

1.1 Motion segmentation

Segmenting individual objects from backgrounds is one of the most important of computer vision tasks. An important clue is provided by motion; humans can easily discern independently moving objects by seeing their motions without knowing their identities. To solve this problem, Costeira and Kanade [1] presented a segmentation algorithm based on feature tracking. Since then, various extensions and applications have been proposed [2, 3, 4, 10].

Costeira and Kanade [1] attributed their algorithm to the Tomasi-Kanade factorization [14], but the underlying principle is a simple fact of linear algebra: the image motion of points that belong to an object moving rigidly in the scene is constrained to be in a four-dimensional subspace [2, 7].

In an attempt to improve the Costeira-Kanade algorithm, Ichimura [3] introduced the discrimination criterion of Otsu [12]. Kanatani [7] introduced model selection and robust estimation to the Costeira-Kanade algorithm and showed that his method, which he called *subspace separation*, far outperforms the Costeira-Kanade algorithm and Ichimura's method.

1.2 Number of motions

In the subspace separation method of Kanatani [7], the number of independent motions is assumed—usually two (the background and one moving object). It has been reported that estimating the number of motions is more difficult than segmentation itself [1, 2].

Gear [2] attempted a complicated statistical analysis for estimating the number of motions and concluded that individual points were likely to be judged as independently moving. This is natural because saying that each point is moving independently (but coincidentally in unison) is, as long as the judgment is based on statistical likelihood alone, always a more likely interpretation than saying that the motions of different points are constrained

It follows that it is *impossible* to infer the number of motions unless some criterion that favors a small number of motions is introduced. Costeira and Kanade [1] dealt with the problem by introducing ad-hoc thresholds, but no guiding principle exists for setting them. In this paper, we show that the problem can be solved by *model selection* without introducing any empirical thresholds.

There is, however, a subtle problem underneath. How can we "compare" different criteria? Usually, an estimation method is regarded as better if the output is closer to the "true value". But what is the "true value"?

1.3 What is the true number of motions?

Suppose six objects are moving in the scene in three ways, say, objects 1 and 2 are moving rigidly as a whole, and so are objects 3 and 4 and objects 5 and 6. What is the correct answer for the number of motions? Obviously, we expect it to be 3, but if we say that it is 6, what is wrong? After segmenting the points into 6 groups, we can compute their 3-D shapes and motions by applying a structure-from-motion algorithm to each group separately. We may find *a posteriori* that the motion of one group is very similar to another, etc., concluding that there are three motions in total. This 3-D interpretation is correct. Should we assign penalty to this "overestimation"?

By the same token, saying that the number of motions is 5 is also correct, and so are the answers 4 and 3. But the answer 2 is wrong. In other words, the correct answers form a spectrum, the lower bound being 3 and the upper bound being the number of points observed. The surest answer is the upper bound, but our intuition is that the smaller the answer is, the better it is *unless it steps over the lower bound*. In other words, we want to minimize the overestimation, but at the same time some "safety margin" is desired. We

must take this into consideration in comparing model selection criteria.

1.4 A posteriori evaluation

Once the number m of motions is given, we can segment the points into m groups by subspace separation [7], but the result may not always be correct. In this paper, we also study the method for evaluating the correctness of the segmentation *a posteriori* without knowing the true answer. Again, model selection plays an important role.

1.5 Organization of this paper

In Sec. 2, we show that the image motion of points that belong to an object moving rigidly in the scene is constrained to be in a four-dimensional subspace and hence estimating the number of motions reduces to estimating the rank of points. In Sec. 3, we summarize standard methods for computing the rank of points. In Sec. 4, we compare different model selection criteria for estimating the rank of a matrix. Sec. 5 describes the standard F-test for testing the correctness of segmentation. Sec. 6 describes the model selection procedure for the same test. In Sec. 7, we show an example using real images. Sec. 8 gives our conclusion.

2. Motion Subspaces

Suppose we track N rigidly moving feature points over M images. Let $(x_{\kappa\alpha}, y_{\kappa\alpha})$ be the image coordinates of the αth point in the κth frame. If we stack the image coordinates over the M frames vertically into a $2M$-dimensional vector in the form

$$\boldsymbol{p}_\alpha = \begin{pmatrix} x_{1\alpha} & y_{1\alpha} & x_{2\alpha} & y_{2\alpha} & \cdots & y_{M\alpha} \end{pmatrix}^\top, \quad (1)$$

the image motion of the αth point is represented by a single point $\boldsymbol{p}_\alpha$ in a $2M$-dimensional space.

We regard the XYZ camera coordinate system as the world coordinate system with the Z-axis taken along the optical axis. We fix an arbitrary object coordinate system to the object and let $\boldsymbol{t}_\kappa$ and $\{\boldsymbol{i}_\kappa, \boldsymbol{j}_\kappa, \boldsymbol{k}_\kappa\}$ be, respectively, its origin and orthonormal basis in the κth frame. Let $(a_\alpha, b_\alpha, c_\alpha)$ be the object coordinates of the αth point. Its position in the κth frame with respect to the world coordinate system is given by

$$\boldsymbol{r}_{\kappa\alpha} = \boldsymbol{t}_\kappa + a_\alpha \boldsymbol{i}_\kappa + b_\alpha \boldsymbol{j}_\kappa + c_\alpha \boldsymbol{k}_\kappa. \quad (2)$$

If we assume orthographic projection, we have

$$\begin{pmatrix} x_{\kappa\alpha} \\ y_{\kappa\alpha} \end{pmatrix} = \tilde{\boldsymbol{t}}_\kappa + a_\alpha \tilde{\boldsymbol{i}}_\kappa + b_\alpha \tilde{\boldsymbol{j}}_\kappa + c_\alpha \tilde{\boldsymbol{k}}_\kappa, \quad (3)$$

where $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ are the 2-dimensional vectors obtained from $\boldsymbol{t}_\kappa$, $\boldsymbol{i}_\kappa$, $\boldsymbol{j}_\kappa$, and $\boldsymbol{k}_\kappa$, respectively, by chopping the third components. If we stack the vectors $\tilde{\boldsymbol{t}}_\kappa$, $\tilde{\boldsymbol{i}}_\kappa$, $\tilde{\boldsymbol{j}}_\kappa$, and $\tilde{\boldsymbol{k}}_\kappa$ over the M frames vertically into $2M$-dimensional vectors $\boldsymbol{m}_0$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$, respectively, the vector $\boldsymbol{p}_\alpha$ has the form

$$\boldsymbol{p}_\alpha = \boldsymbol{m}_0 + a_\alpha \boldsymbol{m}_1 + b_\alpha \boldsymbol{m}_2 + c_\alpha \boldsymbol{m}_3. \quad (4)$$

Thus, the N points $\{\boldsymbol{p}_\alpha\}$ belong to the 4-dimensional subspace spanned by the vectors $\{\boldsymbol{m}_0, \boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$. This fact holds for all affine camera models including weak perspective and paraperspective [13].

It follows that if there are m independent motions, the points $\{\boldsymbol{p}_\alpha\}$ are constrained to be in a $4m$-dimensional subspace of $\mathcal{R}^n$ $(n = 2M)$. Hence, estimating the number of motions reduces to estimating the *rank* of a set of vectors, i.e., the dimension of the subspace they span.

We also observe that discerning m independent motions requires $2m$ or more images. If we are observing 2-D rigid motions within the image plane, the vector $\boldsymbol{m}_3$ in eq. (3) is identically $\boldsymbol{0}$. Hence, the rank of $\{\boldsymbol{p}_\alpha\}$ is $3m$ for m independent motions; we need $1.5m$ or more images for discerning them.

3. Rank Estimation

Mathematically, there are basically three ways for computing the rank r of a set $\{\boldsymbol{p}_\alpha\}$ of N n-dimensional vectors.

1. Define the $n \times n$ *moment matrix* $\boldsymbol{M}$ by

$$\boldsymbol{M} = \sum_{\alpha=1}^{N} \boldsymbol{p}_\alpha \boldsymbol{p}_\alpha^\top. \quad (5)$$

   Let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues, and $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_n\}$ the corresponding orthonormal set of eigenvectors. The rank r equals the number of positive eigenvalues, i.e., the rank of $\boldsymbol{M}$.
2. Define the $N \times N$ *metric matrix* $\boldsymbol{G} = (G_{\alpha\beta})$ by

$$G_{\alpha\beta} = (\boldsymbol{p}_\alpha, \boldsymbol{p}_\beta). \quad (6)$$

   Throughout this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Let $\lambda_1 \geq \cdots \geq \lambda_N$ be its eigenvalues, and $\{\boldsymbol{v}_1, ..., \boldsymbol{v}_N\}$, the corresponding orthonormal set of eigenvectors. The rank r equals the number of positive eigenvalue, i.e., the rank of $\boldsymbol{G}$.
3. Define the $n \times N$ *observation matrix* $\boldsymbol{W}$ by

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_N \end{pmatrix}. \quad (7)$$

   If $n > N$, let its singular value decomposition be

$$\boldsymbol{W} = \boldsymbol{U}_{n\times N} \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_N) \boldsymbol{V}_{N\times N}^\top, \quad (8)$$

   where $\boldsymbol{U}_{n\times N}$ and $\boldsymbol{V}_{N\times N}$ are, respectively, $n \times N$ and $N \times N$ matrices having orthonormal columns. The symbol $\mathrm{diag}(\cdots)$ designates the diagonal matrix with $\cdots$ as its diagonal elements in that order.

If $N \geq n$, let the singular value decomposition of $\boldsymbol{W}^\top$ be

$$\boldsymbol{W}^\top = \boldsymbol{V}_{N\times n}\mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_n)\boldsymbol{U}^\top_{n\times n}, \qquad (9)$$

where $\boldsymbol{V}_{N\times n}$ and $\boldsymbol{U}_{n\times n}$ are, respectively, $N \times n$ and $n \times n$ matrices having orthonormal columns. In either case, the rank r equals the number of positive singular values, i.e., the rank of $\boldsymbol{W}$ and $\boldsymbol{W}^\top$.

However, *all these are in theory only*: they can be applied only when no noise exists in the data and the computation can be done with infinite accuracy. In the presence of noise and with finite computation accuracy, all eigenvalues and singular values are nonzero in general. Hence, we need to truncate small eigenvalues and singular values, but it is difficult to set an appropriate threshold.

4. Model Selection for Rank Estimation

A naive idea for estimating the rank of a set $\{\boldsymbol{p}_\alpha\}$ of N n-dimensional vectors is to fit subspaces of different dimensions to them and to adopt the dimension of the subspace having the smallest *residual*, i.e., the sum of the square distances of the points $\{\boldsymbol{p}_\alpha\}$ in $\mathcal{R}^n$ from the fitted subspace. This does not work, however, because the residual becomes smaller as we fit a higher dimensional subspace. In particular, the residual is zero if we fit the n-dimensional subspace (= the entire space). Also, the residual becomes smaller if the subspace to fit has more parameters to adjust.

Thus, we need to balance the residual against the dimension and the degree of freedom of the subspace. Typical criteria that take this into account are the *geometric AIC* [5, 6] and the *geometric MDL* [9, 11].

4.1 Geometric AIC and geometric MDL

We assume that each image coordinate of the feature points undergoes independent Gaussian noise of mean 0 and a constant variance ϵ^2 at each frame. The degree of freedom of an r-dimensional subspace in $\mathcal{R}^n$ is[1] $r(n-r)$. Hence, the geometric AIC and the geometric MDL are respectively given as follows (see [8] for the details):

$$\begin{aligned} \text{G-AIC} &= \hat{J}_r + 2r(m+n-r)\epsilon^2, \\ \text{G-MDL} &= \hat{J}_r - r(m+n-r)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \qquad (10)$$

Here, L is a reference length for the data. We can use for it an arbitrary value whose order is approximately the same as the data, say the image size; the model selection result is not much affected as long as it has the same order of magnitude.

[1] An r-dimensional subspace of $\mathcal{R}^n$ is specified by r points in $\mathcal{R}^n$, but they can move freely within the subspace. So, the degree of freedom is $rn - r^2$.

Let $\nu = \min(n, m)$. The residual $\hat{J}_r$ is given by

$$\hat{J}_r = \sum_{i=r+1}^{\nu} \sigma_i^2, \qquad (11)$$

where $\{\sigma_i\}$ are the singular values, in descending order, of the observation matrix $\boldsymbol{W}$. Evaluating eqs. (10) for $r = 1, 2, ...$, we choose the value r that minimizes them.

If the noise variance ϵ^2 is not known, we need to estimate it. If an upper bound $r_{\max}$ on the rank r is known, we have the following estimate [5]:

$$\hat{\epsilon}^2 = \frac{\hat{J}_{r_{\max}}}{(n - r_{\max})(N - r_{\max})}. \qquad (12)$$

This can be intuitively understood as follows. If we fit an $r_{\max}$-dimensional subspace, the codimension (= the dimension of the orthogonal directions to it) is $n - r_{\max}$. Hence, the sum of square distances $\hat{J}_{r_{\max}}$ of the N points from it should have expectation $(n - r_{\max})N\epsilon^2$. However, the subspace is fitted so as to minimize $\hat{J}_{r_{\max}}$ by adjusting its $r_{\max}(n - r_{\max})$ degrees of freedom. Hence, the expectation of $\hat{J}_{r_{\max}}$ reduces to $(n - r_{\max})N\epsilon^2 - r_{\max}(n - r_{\max})\epsilon^2 = (n - r_{\max})(N - r_{\max})\epsilon^2$.

4.2 Otsu-Ichimura criterion

The geometric AIC and the geometric MDL effectively truncate the eigenvalues and singular values without using any threshold. A well known automatic thresholding scheme is the discrimination criterion of Otsu [12], which Ichimura [3] used for thresholding matrix elements for motion segmentation using the Costeira-Kanade algorithm. If this is applied to thresholding the singular values $\{\sigma_i\}$ of the observation matrix $\boldsymbol{W}$ given by eq. (7) or (8), we obtain the following criterion, which we call the *Otsu-Ichimura criterion*:

$$\text{OIC} = \frac{r(\nu - r)(\mu_1 - \mu_2)^2}{\sum_{i=1}^{r}(\sigma_i - \mu_1)^2 + \sum_{i=r+1}^{\nu}(\sigma_i - \mu_2)^2}. \qquad (13)$$

Here, we define

$$\mu_1 = \frac{1}{r}\sum_{i=1}^{r}\sigma_i, \qquad \mu_2 = \frac{1}{\nu - r}\sum_{i=1+r}^{\nu}\sigma_i. \qquad (14)$$

The number r that maximizes eq. (13) is chosen as the number of the nonzero singular values.

4.3 Rank estimation experiment

We defined a 10×20 matrix with random elements uniformly generated over $[-1, 1]$. We computed its singular value decomposition in the form $\boldsymbol{V}\mathrm{diag}(\sigma_1, ..., \sigma_{10})\boldsymbol{U}^\top$, the singular values $\sigma_1, ..., \sigma_5$ being, respectively, 3.81, 3.58, 3.09, 2.98, 2.75. Then, we defined the matrix

$$\boldsymbol{A} = \boldsymbol{V}\mathrm{diag}(\sigma_1, ..., \sigma_5, \gamma\sigma_5, 0, ..., 0)\boldsymbol{U}^\top. \qquad (15)$$

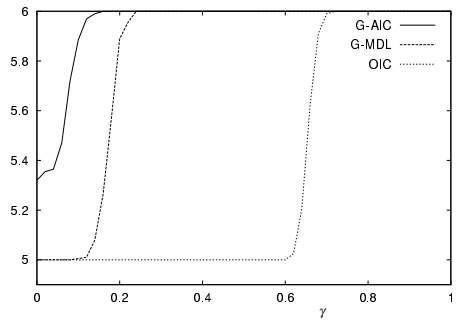

Figure 1: Average estimated rank. The solid lines are for the geometric AIC; the broken lines are for the geometric MDL; the dotted lines are for the OIC.

We added Gaussian noise of mean 0 and standard deviation 0.05 to each element of $\boldsymbol{A}$ independently and estimated its rank with $r_{\max} = 6$. Fig. 1 plots the average number of the rank over 200 trials for each γ. We used the reference length $L = 1$.

The geometric AIC predicts the rank to be 6 with some probability even when the true rank is 5 ($\gamma = 0$), and it mostly predicts the rank to be 6 even for a small value of γ. The geometric MDL almost always guesses the rank to be 5 when the true rank is 5 ($\gamma = 0$), but it keeps guessing the rank to be 5 for a wide range of γ for which the true rank is 6. The OIC is likely to threshold smaller singular values to zero and predict the rank to be 5 up to a fairly large value of γ.

Thus, the geometric AIC tends to be faithful to noise and overestimate the rank, while the geometric MDL tends to ignore noise and underestimate the rank. Their contrasting behavior like this can be seen in other applications, too [9, 11].

In doing our experiment, we have found that if we use a large upper bound for $r_{\max}$, the residual $\hat{J}_{r_{\max}}$ in eq. (12) becomes so small that the lack of significance digits causes underestimation of $\hat{\epsilon}^2$, often regarding it as zero. Hence, the upper bound $r_{\max}$ should be taken to be as small as possible.

4.4 Motion number estimation experiment

Fig. 2 shows five simulated images of 20 background points, 9 object points, and another 9 object points, each independently moving rigidly within the image plane. The object points are linked by edges for the ease of visulalization. The image size is supposed to be 512×512. We added Gaussian noise of mean 0 and standard deviation ϵ to the coordinates of the 38 points independently and estimated the number of independent motions.

Fig. 3 plots the average number of detected motions over 500 trials for each ϵ with the upper bound $m_{\max} = 4$. We used the reference length $L = 600$. The geometric AIC (solid line) tends to overestimate the number of motions but is stable. The geometric MDL (broken line) estimates the correct number of motions when the noise is very small but tends to underestimate it as the noise increases. The OIC (dotted lines) always estimates the number of motions to be 1.

The irregular behavior at the left end of the plot is due to the lack of significant digits for computing too small residuals.

5. F Test for Subspace Separation

Even if a segmentation technique is proved to have high performance by simulations, there is no guarantee that a particular segmentation is correct. So, we need some means for an a posteriori evaluation.

Suppose the N points $\{\boldsymbol{p}_\alpha\}$ that represent the feature motion history are segmented into m groups having N_i points, $i = 1, ..., m$. Let $\hat{J}_i$ be the residual of fitting a d-dimensional subspace $\mathcal{L}_i$ to the ith group. The subspace $\mathcal{L}_i$ has codimension $n - d$ and $d(n - d)$ degrees of freedom. Hence, $\hat{J}_i/\epsilon^2$ should be subject to a χ^2 distribution with

$$\phi_i = (n - d)N_i - d(n - d) = (n - d)(N_i - d) \quad (16)$$

degrees of freedom [5].

If we let $\hat{J}_t$ be the residual of fitting an md-dimensional subspace $\mathcal{L}_t$ to the entire N points $\{\boldsymbol{p}_\alpha\}$, $\hat{J}_t/\epsilon^2$ should be subject to a χ^2 distribution with

$$\phi_t = \sum_{i=1}^{m}(n-md)N_i - md(n-md) = (n-md)(N-md) \quad (17)$$

degrees of freedom *irrespective of the correctness of the segmentation*.

The residual $\hat{J}_i$ is the sum of square distances of the points of the ith group from the subspace $\mathcal{L}_i$, which is the sum of their distances from the subspace $\mathcal{L}_t$ and the distances of their projections onto $\mathcal{L}_t$ from the subspace $\mathcal{L}_i$ (Fig. 4). Let us call the former the *external distances*, and the latter the *internal distances*. The sum of the square internal distances for all the points is $\sum_{i=1}^{m} \hat{J}_i - \hat{J}_t$; the sum of the square external distances is $\hat{J}_t$.

If this segmentation is correct (the *null hypothesis*), $(\sum_{i=1}^{m} \hat{J}_i - \hat{J}_t)/\epsilon^2$ should also be subject to a χ^2 distribution. The noise that contributes to the internal distances and the noise that contributes to the external distances are *orthogonal* to each other and hence independent. So, $(\sum_{i=1}^{m} \hat{J}_i - \hat{J}_t)/\epsilon^2$ has

$$\sum_{i=1}^{m} \phi_i - \phi_t = (m - 1)d(N - md) \quad (18)$$

degrees of freedom [5]. It follows that

$$F = \frac{(\sum_{i=1}^{m} \hat{J}_i - \hat{J}_t)/(m - 1)d(N - md)}{\hat{J}_t/(n - md)(N - md)} \quad (19)$$

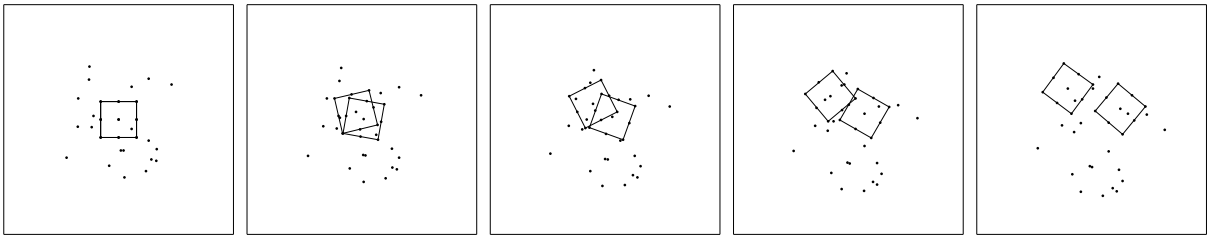

Figure 2: Points moving two-dimensionally.

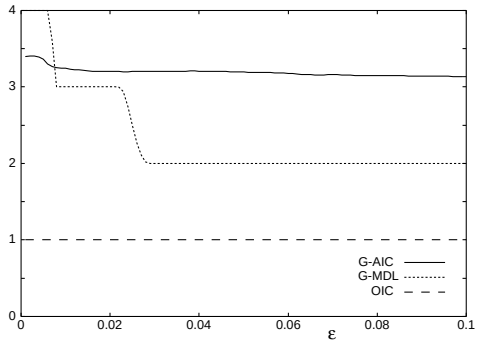

Figure 3: Average estimate number of motions. The solid lines are for the geometric AIC; the broken lines are for the geometric MDL; the dotted lines are for the OIC.

should be subject to an F distribution with $(m-1)d(N-md)$ and $(n-md)(N-md)$ degrees of freedom. If this segmentation is not correct (the *alternative hypothesis*), the internal distances will increase on average while the external distances are not affected. It follows that if

$$F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha) < F, \tag{20}$$

this segmentation is rejected with significance level α, where $F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha)$ is the upper $\alpha\%$ percentail of the F distribution with $(m-1)d(N-md)$ and $(n-md)(N-md)$ degrees of freedom.

6. Model Selection for Segmentation

The result of an F test depends on the significance level, which we can arbitrarily set. The use of the geometric AIC and the geometric MDL can dispense with any thresholds.

The geometric AIC and the geometric MDL for the model that the segmentation is correct are, respectively, $\sum_{i=1}^{m}$ G-AIC$_i$ and $\sum_{i=1}^{m}$ G-MDL$_i$, where G-AIC$_i$ and G-MDL$_i$ are the the geometric AIC and the geometric MDL of the ith group given as follows:

$$\begin{aligned} \text{G-AIC}_i &= \hat{J}_i + 2\Big(dN_i + d(n-d)\Big)\epsilon^2, \\ \text{G-MDL}_i &= \hat{J}_i - \Big(dN_i + d(n-d)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{21}$$

The geometric AIC and the geometric MDL for the model that the points $\{\boldsymbol{p}_\alpha\}$ can be somehow segmented into m motions are given as follows:

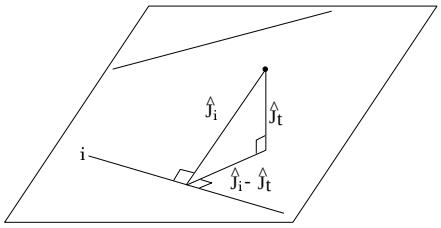

Figure 4: The residual of subspace fitting.

$$\begin{aligned} \text{G-AIC}_t &= \hat{J}_t + 2\Big(mdN_i + md(n-md)\Big)\epsilon^2, \\ \text{G-MDL}_t &= \hat{J}_t - \Big(mdN_i + md(n-md)\Big)\epsilon^2 \log\Big(\frac{\epsilon}{L}\Big)^2. \end{aligned} \tag{22}$$

Since the latter model is correct irrespective of the segmentation result, we can estimate the square noise level ϵ^2 from it as follows:

$$\hat{\epsilon}^2 = \frac{\hat{J}_t}{(n-md)(N-md)}. \tag{23}$$

The condition that the segmentation is not correct is given by G-AIC$_t$ $< \sum_{i=1}^{m}$ G-AIC$_i$ or G-MDL$_t$ $< \sum_{i=1}^{m}$ G-MDL$_i$, which are rewritten, respectively, as

$$2 < F, \qquad -\log\Big(\frac{\epsilon}{L}\Big)^2 < F, \tag{24}$$

where F is the F statistic given by eq. (19). Thus, model selection using the geometric AIC and the geometric MDL has the same form as the standard F test, the only difference being that *the threshold* $F_{(n-md)(N-md)}^{(m-1)d(N-md)}(\alpha)$ *is given automatically without specifying any significance level.* When the noise is small, $-\log(\epsilon/L)^2$ is usually larger than 2. This implies that the geometric AIC is more conservative than the geometric MDL, which is more confident of the particular result.

7. Real image experiment

Fig. 5 shows a sequence of perspectively projected images (above) and manually selected feature points from them (below). Three objects are fixed in the scene, moving rigidly with the scene, while one object is moving relative to the scene. The image size

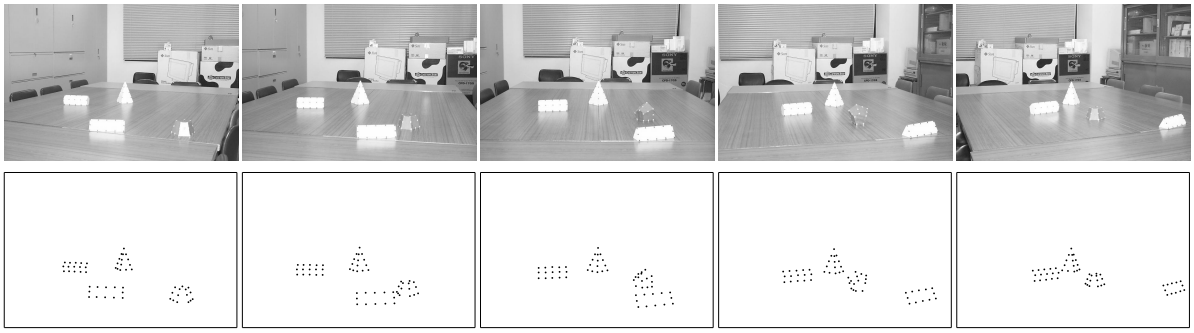

Figure 5: Real images of moving objects (above) and the selected feature points (below).

is 512×768 pixels. Letting[2] $m_{\max} = 2$, we found that the number of independent motions was 2 according to both the geometric AIC and the geometric MDL but was 1 according to the OIC. By subspace separation [7] with $m = 2$, we obtained a correct segmentation, detecting the independently moving object correctly.

The five images shown here were decimated from a motion sequence. Adding five intermediate images, we estimated the number of motions from the ten images with $m_{\max} = 3$, 4, using the geometric AIC and the geometric MDL. In all cases, the computed number of motions was $m = 3$. However, the OIC always estimated the number of motions to be $m = 1$.

Applying subspace separation with $m = 3$, we detected the independently moving object correctly. In this sense, the estimation by the geometric AIC and the geometric MDL is correct, while the estimation by the OIC cannot be correct (see Sec. 1.3).

We then evaluated the reliability of the segmentation. The F statistic of eq. (19) was computed to be $F = 0.893$. The upper 5% percentile is 1.346 in this case, so the correctness of the segmentation is not rejected with significance level 5%. The geometric AIC also selects this segmentation. Using the reference length $L = 600$, we have $-\log(\epsilon/L)^2 = 13.5$, so the geometric MDL also selects this segmentation, but it allows a much wider margin than the geometric AIC.

8. Concluding Remarks

In this paper, we studied the problem of estimating the number of independent motions for segmentation based on feature tracking. We elucidated the mathematical structure of the problem and presented an estimation method using model selection by the geometric AIC, the geometric MDL, and the OIC.

From our experiments, we conclude that the geometric AIC sometimes overestimates the number of motions but this does not affect the correctness of the segmentation, while the geometric MDL sometimes underestimates the number of motions. The OIC always returns an incorrect answer. So, the geometric AIC seems to be a practical choice.

Then, we presented techniques for evaluating the reliability of segmentation *a posteriori*, using the standard F test and model selection by the geometric AIC and the geometric MDL. The use of model selection has the advantage that no significance level needs to be assigned. Again, the geometric AIC seems to be a feasible choice.

[2] In order to let $m_{\max} = 3$, we need seven or more images (see Sec. 2).

References

[1] J. P. Costeira and T. Kanade, A multibody factorization method for independently moving objects, *Int. J. Comput. Vision*, **29**-3 (1998), 159–179.

[2] C. W. Gear, Multibody grouping from motion images, *Int. J. Comput. Vision*, **29**-2 (1998), 133–150.

[3] N. Ichimura, Motion segmentation based on factorization method and discriminant criterion, *Proc. 7th Int. Conf. Comput. Vision*, September 1999, Kerkya, Greece, pp. 600–605.

[4] N. Ichimura, Motion segmentation using feature selection and subspace method based on shape space, *Proc. 15th Int. Conf. Pattern. Recogn.*, September 2000, Barcelona, Spain, Vol.3, pp. 858–864

[5] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, The Netherlands, 1996.

[6] K. Kanatani, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.

[7] K. Kanatani, Motion segmentation by subspace separation and model selection, *Proc. 8th Int. Conf. Comput. Vision*. July 2001, Vancouver, Canada, Vol. 2, pp. 301–306.

[8] K. Kanatani, Model selection for geometric inference, *Proc. 5th Asian Conf. Comput. Vision*. January 2002, Melbourne, Australia.

[9] K. Kanatani and C. Matsunaga, Geometric MDL and its media applications, *Proc. 2000 Workshop Information-based Induction Science*, July 2000, Izu, Japan, pp. 45–51.

[10] A. Maki and C. Wiles, Geotensity constraint for 3D surface reconstruction under multiple light sources, *Proc. 6th Euro. Conf. Comput. Vision*, June–July 2000, Dublin, Ireland, Vol.1, pp. 725–741.

[11] C. Matsunaga and K. Kanatani, Calibration of a moving camera using a planar pattern: Optimal computation, reliability evaluation and stabilization by model selection, *Proc. 6th Euro. Conf. Comput. Vision*, June–July, 2000, Dublin, Ireland, Vol. 2, pp. 595–609.

[12] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Sys. Man Cyber.*, **9**-1 (1979), 62–66.

[13] C. J. Poelman and T. Kanade, A paraperspective factorization method for shape and motion recovery, *IEEE Trans. Pat. Anal. Mach. Intell.*, **19**-3 (1997), 206–218.

[14] C. Tomasi and T. Kanade, Shape and motion from image streams under orthography—A factorization method, *Int. J. Comput. Vision*, **9**-2 (1992), 137–154.